art - Support #24494

art::SummedValue combining handles from different runs

05/28/2020 03:42 PM - Gianluca Petrillo

Status:	Closed		**Start date:**	05/28/2020
Priority:	Normal		**Due date:**	
Assignee:	Kyle Knoepfel		**% Done:**	100%
Category:	Infrastructure		**Estimated time:**	0.00 hour
Target version:	3.06.00		**Spent time:**	4.00 hours
Scope:	Internal		**SSI Package:**	
Experiment:	LArSoft			

Description

I have attempted to "manually" aggregate subrun data products with art::SummedValue, and I have encountered an unexpected behaviour.
It appears to me that the aggregation mechanism does not consider which run the values cover, but only the subrun.
So it happened that in feeding handles from data products on a new run, the object threw an exception of overlapping ranges.
The code that exercises that is a LArSoft module in larsim:version:v08_26_00 for larsoft:version:v08_54_00, module POTaggregator (larsim/EventGenerator).
The code left me the impression that art::RangeSet is designed to cover a single run at a time (things like art::RangeSet::set_run).
I don't know if that is intentional and by design.

Question is whether this is any surprise to you, and whether there is a supported *art* utility to aggregate data products with the proper range checks.

Steps to reproduce (tested on dunegpvm04.fnal.gov:
1. copy all the attached (ICARUS) files on an area with LArSoft CVMFS available.
2.

```
source /cvmfs/larsoft.opensciencegrid.org/products/setup     # set up
setup larsoft v08_53_00 -q e19:prof
xargs file_info_dumper --range-of-validity -- < files.list  # verify that the ranges do not overlap
lar -c POTsummary.fcl -S files.list
```

Output:

```
$ source /cvmfs/larsoft.opensciencegrid.org/products/setup  # set up
$ setup larsoft v08_53_00 -q e19:prof
$ xargs file_info_dumper --range-of-validity -- < files.list  # verify that the ranges do not overlap
==============================
File: simulation_genie_icarus_bnb_20200528T201822-GenGenie.root
Representation: full (default)
------------------------------
 Run: 2
  SubRun: 0 Event range: [1,6)

==============================
File: simulation_genie_icarus_bnb_20200528T201826-GenGenie.root
Representation: full (default)
------------------------------
 Run: 3
  SubRun: 0 Event range: [1,6)

$ lar -c POTsummary.fcl -S files.list
%MSG-i MF_INIT_OK:  Early 28-May-2020 15:35:13 CDT JobSetup
Messagelogger initialization complete.
%MSG
28-May-2020 15:35:14 CDT  Initiating request to open input file "simulation_genie_icarus_bnb_20200528T201822-GenGenie.root"
```

```
28-May-2020 15:35:15 CDT  Opened input file "simulation_genie_icarus_bnb_20200528T201822-GenGenie.
root"
Begin processing the 1st record. run: 2 subRun: 0 event: 1 at 28-May-2020 15:35:15 CDT
Begin processing the 2nd record. run: 2 subRun: 0 event: 2 at 28-May-2020 15:35:15 CDT
Begin processing the 3rd record. run: 2 subRun: 0 event: 3 at 28-May-2020 15:35:15 CDT
Begin processing the 4th record. run: 2 subRun: 0 event: 4 at 28-May-2020 15:35:15 CDT
Begin processing the 5th record. run: 2 subRun: 0 event: 5 at 28-May-2020 15:35:15 CDT
28-May-2020 15:35:15 CDT  Closed input file "simulation_genie_icarus_bnb_20200528T201822-GenGenie.
root"
28-May-2020 15:35:15 CDT  Initiating request to open input file "simulation_genie_icarus_bnb_20200
528T201826-GenGenie.root"
28-May-2020 15:35:15 CDT  Opened input file "simulation_genie_icarus_bnb_20200528T201826-GenGenie.
root"
Begin processing the 6th record. run: 3 subRun: 0 event: 1 at 28-May-2020 15:35:15 CDT
Begin processing the 7th record. run: 3 subRun: 0 event: 2 at 28-May-2020 15:35:15 CDT
Begin processing the 8th record. run: 3 subRun: 0 event: 3 at 28-May-2020 15:35:15 CDT
Begin processing the 9th record. run: 3 subRun: 0 event: 4 at 28-May-2020 15:35:15 CDT
Begin processing the 10th record. run: 3 subRun: 0 event: 5 at 28-May-2020 15:35:15 CDT
28-May-2020 15:35:15 CDT  Closed input file "simulation_genie_icarus_bnb_20200528T201826-GenGenie.
root"
POT from 2 runs:
 run: 2 (1 subruns): 6936326962887897.000000 good POT ( 0 spills); total: 6936326962887897.000000
POT ( 0 spills)
 run: 3 (1 subruns): 6405521164478414.000000 good POT ( 0 spills); total: 6405521164478414.000000
POT ( 0 spills)
Aggregated POT from 2 runs (2 subruns): 13341848127366312.000000 good POT ( 0 spills); total: 1334
1848127366312.000000 POT ( 0 spills)


TrigReport ---------- Event  Summary ------------
TrigReport Events total = 10 passed = 10 failed = 0


TrigReport ------ Modules in End-Path: end_path ------------
TrigReport  Trig Bit#         Run    Success       Error Name
TrigReport     0    0          10         10           0 potsummary


TimeReport ---------- Time  Summary ---[sec]----
TimeReport CPU = 0.013907 Real = 0.021581


MemReport  ---------- Memory  Summary ---[base-10 MB]----
MemReport  VmPeak = 529.334 VmHWM = 125.964


%MSG-s ArtException:  PostEndJob 28-May-2020 15:35:15 CDT ModuleEndJob
---- EventProcessorFailure BEGIN
  EventProcessor: an exception occurred during current event processing
  ---- EventProcessorFailure BEGIN
    EndPathExecutor: an exception occurred during current event processing
    ---- ScheduleExecutionFailure BEGIN
      Path: ProcessingStopped.
      ---- EventRangeOverlap BEGIN
        Attempt to merge event ranges that both contain one or more of the same events
         Run: 2
          SubRun: 0 Event range: [1,6)  vs.
          SubRun: 0 Event range: [1,6)The above exception was thrown while processing module POTac
cumulator/potsummary run: 3 subRun: 0
      ---- EventRangeOverlap END
      Exception going through path end_path
    ---- ScheduleExecutionFailure END
  ---- EventProcessorFailure END
---- EventProcessorFailure END
%MSG
Art has completed and will exit with status 1.
```

History

#1 - 06/01/2020 10:32 AM - Kyle Knoepfel

- Assignee set to Kyle Knoepfel

- Status changed from New to Assigned

- Subject changed from art:SummedValue combining handles from different rus to art::SummedValue combining handles from different runs

We will investigate.

#2 - 06/09/2020 07:56 AM - Kyle Knoepfel

- Target version set to 3.06.00

#3 - 06/09/2020 11:51 AM - Kyle Knoepfel

- % Done changed from 0 to 100

- Status changed from Assigned to Resolved

- Category set to Infrastructure

- Experiment LArSoft added

- Experiment deleted (-)

The intention of SummedValue is that it represents a value corresponding to a particular *range of validity*. According to *art*, a range of validity, as represented by the RangeSet C++ type, contains a single run number and a set of event ranges within that run. For this reason, the SummedValue is not intended to span multiple runs.

We have added the necessary checks so that an exception will be thrown if a multi-run, summed-value operation is attempted.

Implemented with commit <u>art:751a70f</u>.

#4 - 07/10/2020 01:59 PM - Kyle Knoepfel

- Status changed from Resolved to Closed

Files

files.list	116 Bytes	05/28/2020	Gianluca Petrillo
simulation_genie_icarus_bnb_20200528T201822-GenGenie.root	143 KB	05/28/2020	Gianluca Petrillo
simulation_genie_icarus_bnb_20200528T201826-GenGenie.root	136 KB	05/28/2020	Gianluca Petrillo